(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 887 759 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
30.12.1998 Bulletin 1998/53

(51) Int. Cl.[6]: **G06F 19/00**

(21) Application number: **97304388.8**

(22) Date of filing: **23.06.1997**

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
SI

(71) Applicants:
- **SMITHKLINE BEECHAM CORPORATION**
  Philadelphia Pennsylvania 19103 (US)
- **Diversified Pharmaceuticals Services**
  South Edina, MN 55435 (US)

(72) Inventors:
- **Wong, Bruce**
  Philadelphia, Pennsylvania 19103 (US)
- **Friedman, Felix**
  France Avenue, South Edina, MN 55435 (US)

(74) Representative:
Giddings, Peter John, Dr. et al
SmithKline Beecham plc
Corporate Intellectual Property,
Two New Horizons Court
Brentford, Middlesex TW8 9EP (GB)

(54) Method and system for identifying at risk patients diagnosed with congestive heart failure

(57) A computer-implemented technique including database processing is used to identify at risk congestive heart failure patients where information about patients exists in a claims database. The technique includes processing the patient information in the claims database to find and extract claims information for a group of congestive heart failure patients. Next, using the extracted information, a set of events, relevant to congestive heart failure, is defined. Next, the extracted information and set of events are processed to create event level information which is organized with respect to events rather than claims. A time window is defined to provide a timeframe from which to judge whether events should be considered in subsequent processing; and, a set of variables is defined as being potential predictors of adverse health outcomes. Subsequently, the event level information, using the time window and the set of variables, is processed to generate an analysis file. Statistical analysis, such as logistic regression, is performed on the analysis file to generate a prediction model where the prediction model is a function of a subset of the set of variables. Finally, the prediction model is applied to the same or another claims database to identify and output at risk patients, diagnosed with congestive heart failure, likely to have adverse health outcomes.

EP 0 887 759 A1

Description

SCOPE OF THE INVENTION

This invention relates to database processing techniques and, more particularly, it relates to identification of congestive heart failure patients having a high risk of adverse health outcomes using various database processing techniques.

BACKGROUND OF THE INVENTION

Congestive heat failure is the symptomatic description of end organ failure of the heart muscle to act as a pump. The failure of the heart to pump blood leads to accumulation of fluid or congestion in the body areas from which the blood was to be pumped. In the case of right ventricular failure, the consequence is systemic edema often manifest initially by symptomatic dependent edema at peripheral sites (e.g. ankles). In the case of left ventricular failure, the fluid accumulation occurs in the lungs and is manifest by pulmonary edema, the primary symptom of which is shortness of breath. Left heart failure is more common and important because of its relative size and the physiologic function of providing systemic circulation to critical organs.

The primary underlying cause of heart failure (low output failure) is coronary artery disease leading to myocardial infarction and cardiac muscle death. Additional pathologies include cardiomyopathies and myocarditis, valvular dysfunction, metabolic and endocrine abnormalities (e.g., hypothyroidism, alcohol, Mg deficiency etc.) pericardial abnormalities.

In 1992, Conjestive Heart Failure (CHF) prevalence was estimated to be approximately 1% of the general population, 6.8% of patients over the age of 65, and 0.2%-0.4% of the managed care population (prior to the active recruiting of senior members). In absolute numbers, this represents approximately 2.76 million patients in the USA. Subclassification by severity provides estimates of 10% as severe, 40% moderate, and the remainder mild. New cases of diagnosed CHF are expected to increase 29%, from 495,7000 per year in 1992 to 642,000 by the year 2007. This increase can be attributed to several factors, including the general aging of the population, as well as improvements in physician awareness and diagnostic techniques.

There has been a marked increase in the number of seniors enrolling in managed care plans (MCOs). In some plans, seniors can account for up to 50% of total enrollment. In addition, the number of seniors enrolled in Medicare risk programs is projected to grow by 28% to 2.3 million by the end of 1995. Growth in Medicare risk programs is expected to continue, due to the anticipated passage of the Medicare reform bill, which will encourage seniors to join health plans. Also, the current "50/50" provision that forces MCOs to enroll one non-Medicare patient for every Medicare patient enrolled may be overturned, which will make it easy for plans to concentrate on Medicare patients.

Patients with CHF are living longer as treatment options increase. This means that MCOs will need to be able to cost-effectively manage CHF for an ever-growing number of patients. This is made apparent by several other facts. CHF was the most expensive health care problem according to the Healthcare Financing Agency (HCFA) in 1992. Approximately 5.45 billion dollars were spent on hospitalization, or 4.8% of the total Diagnositc Related Group (DRG) budget. CHF is also the single most frequent cause of hospitalization for people age 65 and older. As was previously noted, it is projected that one percent of the general population has CHF, which translates to 6.8% of the elderly. Interestingly, even in MCOs that do not actively recruit seniors, CHF can account for up to 2% of the overall MCO medical loss rate.

Treatment of CHF is primarily symptomatic through three pathophysiologic mechanisms:

1. Afterload reduction - Reduction of the resistance against which the heart must pump blood. Primarily achieved pharmacologically although in severe situations aortic counter pulsation is used.
2. Cardiac Inotropic state - Improvement of the efficiency of the heart through pharmacological agents acting directly on myocardium (e.g. cardiac glycosides), reduction in starling equilibrium's through chamber size reduction and improvement in coronary artery blood flow.
3. Pre-load reduction - Reduction in the volume of blood entering the heart. This can be achieved through pharmacological interventions (diuretics, vasodilators), dietary (salt and water restriction) and physical means (rotating tourniquets, blood volume reduction).

The primary pharmacological interventions, therefore, consist of agents from three primary classes, diuretics, cardiac glycosides and Angiotesin Converting Enzyme (ACE) inhibitors. In addition, recent studies have shown benefit from the use of beta-blockers with anti-oxidant properties. Ultimately, cardiac transplantation now offers an alternative to terminal life threatening heart failure.

The complexity of care required, as outlined above, indicates the need for substantial physician and patient involvement in the optimal management of the condition including the appropriate selection and use of pharmacological inter-

ventions and the modification of individual life styles to achieve best results from interventions.

Obviously, an overall objective of any CHF Disease Management Program should be to improve the quality of treatment and outcomes for patients with CHF while, at the same time, achieving cost savings. An important step in doing so is to identify patients who are at high risk of adverse outcomes and assuring "best practice" treatment of these patients. There is, therefore, a real need to identify patients who are at high risk of hospitalization for CHF.

SUMMARY OF THE INVENTION

The present invention involves a computer-implemented method for generating a model to identify at risk patients diagnosed with congestive heart failure, from information about patients existing in a claims database, said method comprising the steps of 1) processing, based on predetermined criteria, the patient information in the claims database to find and extract claims information for a group of congestive heart failure patients; 2) defining, using the information available in the claims database, events relevant to congestive heart failure; 3) processing the extracted claims information and the defined events to create files containing event level information; 4) defining a time window for providing a timeframe from which to judge whether events should be considered in subsequent processing; 5) defining a set of variables as potential predictors; 6) processing the event level information, using the time window and the set of variables, to generate an analysis file; and 7) performing statistical analysis on the analysis file to generate a prediction model, said prediction model being a function of a subset of the set of variables.

Another aspect of the present invention involves a computer-implemented method for identifying, using the generated model, at risk patients diagnosed with congestive heart failure, said method comprising the additional step of applying the prediction model to a processed claims database to identify and output a file listing the likelihood of each patient having an adverse health outcome.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is best understood from the following detailed description when read in connection with the accompanying drawing, in which:

Figure 1A is a high-level flowchart illustrating an exemplary overall process of the present invention.

Figure 1B is a high-level flowchart illustrating an exemplary process application for the present invention.

Figure 2 is a high-level block diagram illustrating three exemplary sources of information suitable for use with the present invention.

Figure 3 is a data structure diagram which shows an exemplary format in which the information from the sources of Figure 2 are stored in a research database.

Figure 4 is a data structure diagram which shows an exemplary format for an event level file generated during the process shown in Figure 1.

Figure 5 is a data structure diagram which shows an exemplary format for an analysis file generated, in part, from the event level file shown in Figure 4 and during the process shown in Figure 1.

Figure 6A is a simple bubble diagram illustrating the concept of analysis and prediction zones as used in the present invention.

Figure 6B is a time-line diagram which shows a first exemplary time window scheme suitable for use in processing the data, in part, from the event level files shown in Figure 4.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is embodied in apparatus and a method to identify, in a predetermined population of congestive heart failure patients, those patients at high risk of adverse health outcomes. The identification of these high risk patients being an initial stage in attempts, e.g., targeted interventions, to prevent and/or improve their health outcome.

Initially, one or more sources of information are used which allow for the identification of an initial population of congestive heart failure patients. Examples of sources include health care providers such as doctors, hospitals and pharmacies which all keep records for their patients. The individual records for each of these providers, however, may be scattered, difficult to access, and/or have many different formats.

On the other hand, a more comprehensive source containing this type of information exists in the health care claims records of any given benefits provider.

Turning to the figures, Figure 1A is a high-level flowchart illustrating an exemplary overall process of the present invention. As illustrated in Figure 1, the "raw" claims information is received and stored in a database (e.g., DB2 format) represented by block 110. In the world of claims processing, before this database of "raw" information can be useful, some pre-processing, step 112, is generally performed which may include rejecting claims, reconciling multiple claims

and so on. The output of this preprocessing step, represented by block 114, is a "cleaner" database now stored, in the exemplary embodiment, in SAS format.

SAS is a well known format and software package produced by SAS Institute, Inc. of Cary, North Carolina. It should be noted that other data processing and storage formats, as appreciated by those skilled in the art, could be used in the storage and processing of data.

It should also be noted that SAS formats, programming techniques and functions are more fully described in the SAS/STAT User's Guide, Version 6, Fourth Edition, Volumes 1 and 2, 1990 and the SAS Language: Reference, Version 6, First Edition, 1990 which are both herein incorporated by reference for their teachings regarding the SAS language, SAS programming, functions and formats.

Moreover, the SAS routines used for processing information as part of the present invention are used for computational operations, executed on a computer and stored on a storage medium such as magnetic tape, disk, CD ROM or other suitable medium for purposes of storage and/or transportability. The stored software can then be used with a computer such as a PC.

The claims records of the benefits provider, although containing important information, may not be organized in a manner for efficient analysis. Thus, the next step is to perform another processing step (e.g., screening for congestive heart failure patients, age, etc.), represented by block 116, to transform the "raw" data into a more appropriate and useful database. That is, the output data from the processing (i.e., extraction) step is a subset of the "raw" information and represents an initial universe of congestive heart failure patients upon which further processing is performed.

A next step, which is optional, is to perform a "quality check" on the initial universe of congestive heart failure patients. This step is somewhat subjective. This processing step, represented by block 118, using intermediate output files, performs a refinement of the extracted information by, for example, checking to see if an imbalance exists in the extracted information. This step, essentially a common sense check, can be performed as many times as necessary to ensure the integrity of the database data. At this point, the database data exists at the claim level.

The information existing at the claim level provides various information in the form of raw data elements. From the claims level data, the next processing step, represented by block 120, creates new files (e.g., primary file 1 and primary file 2) by reformatting the information into an event level format.

Before this occurs, a set of events (e.g., doctor visit for congestive heart failure) relevant to congestive heart failure are defined using a combination of both the raw data elements available from the claims information and clinical knowledge about congestive heart failure. With these events defined, the claims level information is used to create new files based on events rather than claims. Having the information in an event level format is an important aspect of the present invention in that, among other things, it allows for added flexibility in subsequent analysis.

As depicted by block 122, further processing is performed on the event level data to generate an analysis file. In particular, the processing is performed using input information representative of a sliding time window and a plurality of variables. The time window input limits the time periods in which the events from the primary files are considered. That is to say, the time window is used to identify an analysis region and a prediction region where activity in the analysis region is used to predict some predetermined outcome in the prediction region. The selection of variables, both dependent and independent, for analysis, is an important step impacting the accuracy of the final prediction model. The dependent variables are representative of the desired result (i.e., an adverse health outcome to be predicted); whereas, the independent variables are representative of predictors. This processing step, step 122, can be easily re-programmed, via the input parameters, for various time window adjustments as well as various variable modifications. The analysis file generated at this step is a member level file which means it is broken down by member.

With the analysis file in hand, a model or technique for identifying high risk subgroups is determined. That is, as represented by step 124, the analysis file is used to develop an identification technique represented by an equation incorporating a subset of the initial variables programmed into the above-mentioned processing step. The resulting subset are those variables which best reflect a correlation to adverse health outcomes, consequently, resulting in substantial use of health care resources (e.g., funds). It should be noted that the determination of the initial as well as the final variables is an important aspect of present invention as the variables may significantly impact the accuracy of the identification of the subgroup.

The above model for identification can be developed, step 124, in various ways using statistical techniques. The technique used in the exemplary embodiment of the present invention for generating the model is multiple logistic regression.

Figure 1B is a high-level flowchart illustrating an exemplary process of the application of the present invention. Having developed the model, as shown in Figure 1A, it can then be applied to updated claims data, step 132, or to other databases of congestive heart failure patients (e.g., claims information for other benefits providers), in order to identify at risk patients diagnosed with congestive heart failure, step 134, allowing for various types of targeted intervention to properly care for suffering patients as well as maximize the effective allocation of health care resources.

Exemplary Embodiment of the Invention

Although the present invention is illustrated and described below with respect to specific examples of a method and system for identifying congestive heart failure patients at high risk for adverse health outcomes, the invention is not intended to be limited to the details shown. Rather, various modifications may be made in the details within the scope and range of equivalents of the claims and without departing from the spirit of the invention.

As mentioned, the present invention is designed to identify patients with congestive heart failure at high risk of adverse health outcomes. The identification of this high risk subgroup being the first step in being able to try different treatment techniques (e.g., targeted interventions).

Initially, a source of information is required which allows for the identification of a population of congestive heart failure patients. A comprehensive source containing this type of information exists in the health care claims records of many benefit providers. As is known, claims for drugs, doctors and hospitals are received and processed for payment/reimbursement. In the exemplary embodiment of the present invention, this claims information is entered into a DB2 database on a benefits provider's computer system (not shown).

Figure 2 is a high-level block diagram illustrating three exemplary sources of information suitable for use with the present invention. As illustrated in Figure 2, the claims information of such a provider would typically include three sources: pharmacy claims (Rx) 210, doctor (DR) claims 212, and hospital (HL) claims 214. As listed on the blocks representing the claims information, many types of information would be available from the respective claims including drug codes, physician's names, diagnosis codes, procedures, various dates and other important information. Much of this information is referenced using codes, such as drug codes, procedure codes and illness codes. Appendices I-XIV provide exemplary listings of various codes used with the present invention. These codes were selected for processing purposes of the present invention from a voluminous source of codes and, as will be appreciated by those skilled in the art, may be modified to include/exclude codes deemed more/less useful at the various stages of processing.

It should also be noted that in the health care industry, such as practiced in the United States of America, various codes are used in claims information for indicating which procedures, treatments, diagnoses, drugs, etc. are being claimed. For the exemplary embodiment of the present invention, the selected codes are shown in Appendices I-XIV. These codes were found in Physician's Current Procedural Terminology (CPT), American Medical Association (1995) and St. Anthony's ICD-9-CM Code Book (1994) which are both incorporated herein by reference for their teaching of codes and sources of codes. As will be appreciated by those skilled in the art, any set of codes, representative of the various procedures, treatments, diagnosis, drugs, etc. relevant for use with the present invention would suffice. Reference to such codes occurs throughout this specification.

The DB2 database represents a source of "raw" data elements which require processing. A first step in processing this raw data is to perform data integrity checks (e.g., rejected or reconciled claims). Subsequently, the data is routinely download into a "research" database. The research database is a claims level database in SAS format.

Exemplary formats, for each of the Rx, DR and HL claims, of the records contained in the research Database, are shown in Figure 3. As shown in Figure 3, claims are listed from claim 1 to claim x and the appropriate information, for the particular service provider (e.g., Rx) being claimed, is also presented.

Once in SAS format, SAS procedures process the information to 1) identify patients with congestive heart failure (step 116), 2) process the claims level information into event level information (step 120), 3) using predetermined variables and timeframe schemes, generate analysis files for analysis purposes (step 122) and 4) create a prediction model as a function of those variables most reflective of the correlation to an adverse health outcome (step 124).

It should be mentioned that, from a statistical perspective, an important consideration in developing prediction models from datasets is sample size. To maximize the integrity of the prediction model, sample size is an important factor. Prevalence of congestive heart failure, as mentioned in the Background section, is reported to be approximately 1%; however, desirable sample sizes which may be used to determine prediction equations depend on the magnitude of association between variables. As these associations are initially unknown, all patients within any individual plan are initially included.

The first step, extracting patients with congestive heart failure (step 116), uses various parameters to define which patients qualify for the overall initial universe of CHF patients to be considered.

For example, in the exemplary embodiment of the present invention, only patients having 1) a continuous enrollment with the benefits provider of 6 months or longer and 2) having his/her first claim for CHF or treatment with CHF-related medication at 30 years old or older are eligible. Of course, these criteria are exemplary and could be modified such that 12 months or 3 months of enrollment is satisfactory or that an individual must be 18 years of age. In the exemplary embodiment of the present invention, the claims extraction step, step 116, extracts all claims data for patients with either an appropriate code for CHF (see Appendix I) or for treatment with an CHF-related drug (see Appendix XIV).

Subsequent to the claim extraction step, the claim adjustment and integrity checks are optionally performed, step 118. To do so, from the dataset defined above, intermediate output files are generated which contain information for processing purposes. In the exemplary embodiment of the present invention, information for the following items are

generated for review to determine if the data within intermediate data files is in general agreement with the common clinical knowledge and experience and with literature evidence as to event frequencies.

A) First, a frequency count of the number of enrollment periods for the members is generated. Then, for members with multiple enrollment periods of at least 6 months duration, it is determined if a CHF diagnosis is present in each enrollment period. Consequently, enrollment periods without a CHF diagnosis are excluded and, for members with multiple enrollment periods that have a CHF diagnosis, only the most recent enrollment period that contains a CHF diagnosis is kept.

B) For the one enrollment period for all remaining members, ALL COSTS encountered by that member during the entire enrollment are identified. A complete proc univariate for ALL COSTS is provided for each plan separately and all plans together. It should be noted that "proc univariate" is a SAS procedure which generates descriptive statistics (e.g., mean, standard deviation, etc.)

C) From the ALL COSTS determined above, costs which are specifically cardiovascular (CV) COSTS are identified. In doing so, a cost is considered to be a CV COST, if a claim from the DR or HL file has any CV ICD-9 code in the first or second position. If a claim from the Rx file is from therapeutic class 04000 then it is counted as a CV claim and count cost as CV cost. A complete proc univariate for CV COSTS is also provided for each plan separately and all plans together.

D) From the CV COSTS, those costs which are specifically CHF COSTS are identified. A cost is considered to be CHF COST if a claim from the DR or HL file has any CHF ICD-9 code in the first or second position. A complete proc univariate for CHF COSTS is also provided for each plan separately and all plans together.

E) For all member enrollment periods remaining, the total member months for each plan is calculated separately and together. In doing so, a member is considered enrolled during any month that they were enrolled for at least one day. For this, a complete proc univariate is provided for member months for each plan separately and all plans together.

F) Finally, a unique member count is provided for all patient status code = 20 within the remaining enrollment periods. It is noted that status code = 20 indicates that the patient has expired or did not recover.

It should be noted that, regarding the cost calculations, the following guidelines apply in the exemplary embodiment of the present invention:

a. the cost of inpatient hospitalizations, emergency services, physician/outpatient, and other medical services on a per claim basis are considered to be:

AMTPAID + AMTCOPAY + AMTRESERVE + AMTDEDUCT

b. the cost of drugs are considered to be:

AMTPAID + AMTCOPAY

It should also be noted that, for purposes of a cost heirachy, the following rules were used in the exemplary embodiment of the present invention.

1. Only hospitalizations for CHF can spawn other events.
2. Hospital costs include all Rx, procedure, physician charges.
3. Hospital visits can generate Rx and procedure events with costs set to zero (included in hospital cost).
4. Hospital visits cannot generate separate doctor visit events.

The above information for use in performing preliminary evaluations as to the integrity of the data is exemplary and could be modified to include/exclude parameters which are shown to be more/less useful within the spirit of the present invention.

With this information, a "quality check" is performed on the initial universe of CHF patients to make sure that the final results, i.e., prediction model, is not unreasonably skewed due to imbalanced input information. This processing step, block 118, using intermediate output files, allows for a refinement of the extracted information by, for example, checking to see if an imbalance exists in the extracted information which may otherwise taint the integrity of a prediction model. Step 118, in the exemplary embodiment, is performed manually by viewing the intermediate output files. It is contemplated, however, that using various threshold values, this information could be automatically processed to flag a potential imbalance.

Having now extracted and refined the claims level information according to various predetermined criteria deemed relevant for subsequent processing purposes, the information is converted into an event level format.

To provide processing flexibility, particularly in assigning time windows for analysis, the above-mentioned second step (i.e., converting the claims level information into event level information, step 122) is employed to generate two primary data files from which an analysis file can be created.

In the exemplary embodiment of the present invention, primary data file 1 is a member level file and contains all data of a static nature (i.e., not time sensitive) such as 1) Member Key, 2) Date of birth, 3) Gender, 4) First available date of enrollment (i.e., start of dataset or enrollment date), 5) End data of enrollment (i.e., end of dataset or last date of enrollment), 6) Date of first CHF diagnosis (ICD-9 code in any position), 7) Date of last CHF diagnosis (ICD-9 code in any position), 8) Date of first CHF hospitalization, 9) Date of last CHF hospitalization, 10) Date of first diabetes event (therapeutic class 081xxx OR ICD-9 code 250xx in any position), 11) Date of first dysrhythmia event (therapeutic class 047xxx OR ICD-9 codes 426xx-427xx in any position), 12) Date of first respiratory event (therapeutic class 151xxx OR ICD-9 codes 490xx-496xx in any position), 13) Date of patient status code=20.

Primary data file 2 is an events level file with a record for each event ordered by member and the chronological date of the event, in the present invention, presented in descending order of event date.

It should be noted that an event, sometimes referred to as an episode, is an occurrence which, based on clinical knowledge, is deemed relevant to CHF. Having knowledge of what raw data elements are available from the claims, a set of events is defined directly or indirectly from the data elements where events can be based on an individual data element, combination of data elements or derived from individual or multiple data elements.

Figure 4 is an exemplary list of events and format for primary file 2 (an event level file).

With respect to Figure 4, in the exemplary embodiment of the present invention, the following exemplary ground rules are established for providing counts for the various events:

I. Count as a HOSPITALIZATION event (using both 1st and 2nd ICD-9 codes) a claim having a from and through date of at least one day AND having a site code of 04. It is noted that a site code distinguishes between the sites at which the service under consideration took place (e.g., emergency room, doctor's office, etc.). It should be noted that costs go to 1st ICD-9 code category only. Also, if a new hospitalization occurs within one day of discharge from a previous hospitalization, the two hospitalizations are bridged into one. If a new hospitalization occurs greater than one day following a previous hospitalization, the second hospitalization is considered a new one.

II. Count as an ER VISIT event (using both 1st and 2nd ICD-9 codes) a claim having a site code of 07, 08 or 10 OR a claim with the following the Hospital Common Procedure Coding System (HCPCS) codes: A0010-A0070, A0215-A0225, A0999 with a provider code = 81. It should be noted that costs go to 1st ICD-9 code category only.

III. Count as an OFFICE VISIT event (using only one ICD-9 code) a claim having a site code of 01 or 06 AND having a unique date of service (DOS) but allow for different provider keys on the same DOS (if same provider key on same DOS, consider to be the same office visit) BUT if an office visit event occurs during a hospitalization, do not generate an office visit event (Attribute all costs for this event to the hospitalization). ALSO count as an OFFICE VISIT a claim with the following HCPCS codes A0080-A0210 with provider code = 81. For all other office visit events, costs go to 1st ICD-9 code category only. It should be noted that the following provider keys are not considered as separate office visits and should be bridged with an office visit that occurs on the same DOS if one exists: 1) 24 (therapeutic radiology), 2) 34, 35 (independent lab), 3) 55 (hosp o/pat lab x-ray).

The Three Event Types above are further defined according to the associated Diagnoses that follow:

Count as a Congestive Heart Failure (CHF) Hospital, Emergency Room (ER) or office visit if the first or second ICD-9 code is from Appendix I. Count as an Ischemic Heart Disease (IHD) Hospital, ER or office visit if (referring to Appendix II) a claim with one of the following ICD-9 codes in the first or second position:

Subclass	Description
1. 410xx	Acute myocardial infarction
2. 411xx	Other acute and subacute forms of ischemic heart disease
3. 412xx	Old myocardial infarction
4. 413xx	Angina pectoris
5. 414xx	Other forms of chronic ischemic heart disease

Count as a Diabetes Hospital, Emergency Room or office visit if the first or second ICD-9 code is from Appendix III. Count as a Dysrhythmia Hospital, ER or office visit if the first or second ICD-9 code is from Appendix IV. Count as a Hypertension Hospital, ER or office visit if the first or second ICD-9 code is from Appendix V. Count as a Lifestyle Hospital, ER or office visit if the first or second ICD-9 code is from Appendix VI. Count as a Other Heart Disease Hospital, ER or office visit if the first or second ICD-9 code is from Appendix VII. Count as a Respiratory Disease Hospital, ER or office visit if the first or second ICD-9 code is from Appendix VIII. Count as a Thyrotoxicosis Hospital, ER or office visit if the first or second ICD-9 code is from Appendix X. Count as a Pulmonary Embolism Hospital, ER or office visit if the first or second ICD-9 code is from Appendix XI. Count as a Anemia Hospital, ER or office visit if the first or second ICD-9 code is from Appendix XII. Count as an Infection Hospital, ER or office visit if the first or second ICD-9 code is from Appendix XIII.

Count as an Other Disease Hospital, ER or office visit if the first or second ICD-9 code is not from any of the above Appendices.

Count as a MISCELLANEOUS MEDICAL EVENT any claim having a site code of 02, 03, 04, or 09 AND any claim that cannot be linked to one of the other type of events (ER, Hospitalital or Office Visit). In the exemplary embodiment of the present invention, subclasses of MISC. MED. EVENT include CV or NON-CV where: Count as a CV Miscellaneous event if the first or second ICD-9 code is from Appendix IX, and count as a Non-CV Miscellaneous event if the first or second ICD-9 code is NOT from Appendix IX.

Additionally, count as a Routine CV Procedure any claim with one of the following codes:

CPT	80162	Digoxin level
CPT	80190	Procainamide level
CPT	80194	Quinidine level

The following procedures are desirably associated with a CV diagnosis from Appendix IX to be considered a Routine CV Procedure:

CPT	80002-80019	Automated blood tests
CPT	93000-93014	ECG related
CPT	93040-93278	ECG related
ICD-9	8951-8953	ECG related
CPT	93307-93350	ECHO related
ICD-9	8968	ECHO related
ICD-9	8972	ECHO related
CPT	71010-71035	Chest radiology
ICD-9	8744-8749	Chest X-ray

Count as an Intermediate CV Procedure any claim with one of the following codes:

CPT	71090	Pacemaker
CPT	93015-93024	Stress Testing
CPT	78460-78489	CV radiology
CPT	78401 78412	(old codes for CV radiology)

(continued)

CPT	93501-93570	Cardiac Cath
ICD-9	3721-3723	Cardiac Cath
CPT	93600-93660	EPS
ICD-9	3726-3727	EPS
CPT	93720-93799	Other

Count as a Critical CV Procedure any claim with one of the following codes:

CPT	92950-92996	CPR, etc.
ICD-9	3791	CPR, etc.
CPT	33960-33978	Cardiac assist

Count as a CV Surgery any claim with one of the following codes:

CPT	33010-33100	Pericardial procedure
CPT	33200-33261	Pacemaker
ICD-9	3964	Pacemaker
CPT	33400-33492	Valve surgery
ICD-9	35xx	Valve surgery
CPT	33510-33560	CABG
ICD-9	361x	CABG
ICD-9	360x	Angioplasty
CPT	33600-33855	Congenital mal. surgery
CPT	33930-33950	Heart (lung transplant
CPT	V421	Heart transplant
ICD-9	375x	Heart transplantation
ICD-9	336x	Combined heat-lung transplantation

It should be noted that the procedure cost is assigned to the ER, physician office visit or hospitalization in which the procedure occurred. If a procedure occurs outside of the assigned site codes for these events, consider the procedure to be a CV Miscellaneous medical event and assign the costs accordingly.

Regarding ACE Inhibitor:

(a) ACE Inhibitor therapeutic class 045400
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # ACE inhibitor Rx's in the month.

Regarding Loop Diuretic:

(a) Loop diuretic therapeutic class 043100
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of Loop diuretic rx's in the month.

Regarding Other Diuretic:

(a) Other diuretic therapeutic class 043200 and 043300
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of other diuretic rx's in the month.

Regarding Digoxin:

(a) Digoxin therapeutic class 041000
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of digoxin rx's in the month.

Regarding Beta Blocker:

(a) Beta Blocker therapeutic class 044000
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of beta blocker rx's in the month.

Regarding Calcium Channel Bockers:

(a) Calicum Channel Blocker therapeutic class 04200
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of calcium channel blocker Rx's in the month.

Regarding Other CV Drug:

(a) Other CV Drug therapeutic class 04000 (but not ther classes 041000, 042000, 043100, 043200,. 043300, 044000, 045400)
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of other CV drug rx's in the month.

Regarding Non-CV Drug:

(a) Other Non- CV drug therapeutic class NOT 04000, 111200, 083000, 134000, OR 137000.
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of non-CV drug rx's in the month.

Regarding Na/H2O retention:

(a) Other Na/H2O retention (NSAID- ther class 111200, Steroids - ther class 083000, estrogens - ther class 134000 and 137000)
(b) Cost = 0 if generated from a hospital claim.
(c) Category indicator = # of Na/H2O rx's in the month.

After generating the two primary files using the above described instructions and rules, corresponding to step 120 of Figure 1, further processing is performed on the event level data to generate an analysis file, step 122. An exemplary format for the analysis file is shown in Figure 5. As shown, the format of the analysis file includes a list of members in a first column of a table. Across the top of the table is a list of variables, described in detail below. And, the body of the table provides indications as to a member's relation to a listed variable.

In particular, the processing from the primary files to the analysis files includes an algorithm defined, in part, by a time window and a plurality of variables. The algorithm can be re-programmed for various time window adjustments as well as variable modifications. The analysis file generated at this step is a member level file (i.e., organized with respect to members). The main analysis files are member level files derived from the information in the primary files.

Each main analysis file is created to take into account a single reference time window of censored events and prediction window of interest for that file. Each new time window applied to the data, in the exemplary embodiment, uses another main analysis file.

To generate the analysis file, a time window scheme, using a plurality of variables, is applied to the event level data. Discussing the variables first, included in the processing are both independent and dependent variables. The inde-

pedent variables represent potential predictors of the adverse health outcomes; whereas, the dependent variables represent the adverse health outcome to be predicted.

To determine exemplary independent variables for step 122, as many of the original data elements as possible are used, assuming nothing about CHF. Then, based on clinical knowledge, additional variables are created. Furthermore, combinations of the data elements and/or variables, based on clinical knowledge, are used as variables. Finally, some variables may be created and used based on their potential utility as leverage points in disease management.

In the exemplary embodiment of the present invention, the plurality of variables, in addition to each of the items in the event file, currently used by step 122 in the SAS routine for generating an analysis file are shown below in Table 1. It is noted that each of the events in Figure 4 are automatically considered an independent variable for processing.

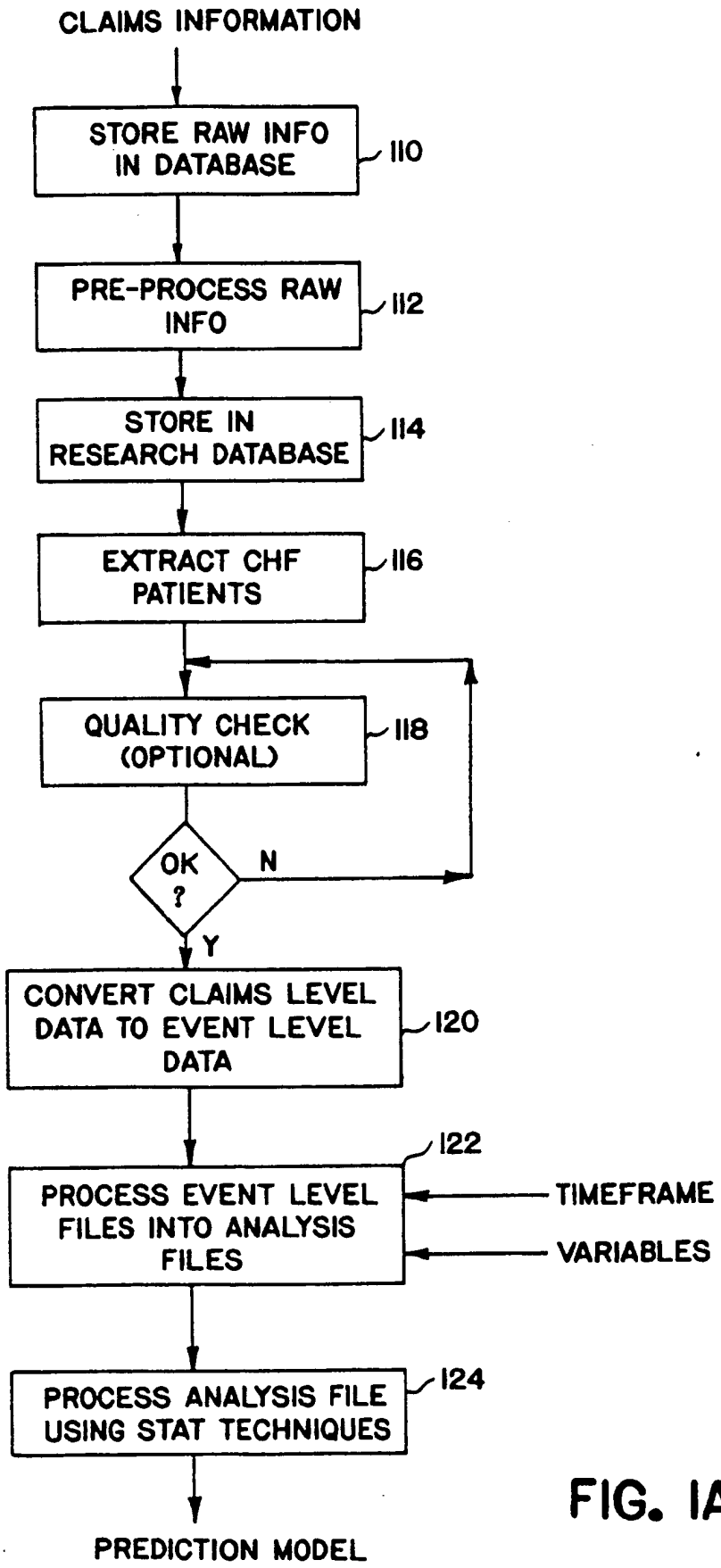
CLAIMS INFORMATION
STORE RAW INFO IN DATABASE
110
PRE-PROCESS RAW INFO
112
STORE IN RESEARCH DATABASE
114
EXTRACT CHF PATIENTS
116
QUALITY CHECK (OPTIONAL)
118
OK ?
N
Y
CONVERT CLAIMS LEVEL DATA TO EVENT LEVEL DATA
120
122
PROCESS EVENT LEVEL FILES INTO ANALYSIS FILES
TIMEFRAME
VARIABLES
PROCESS ANALYSIS FILE USING STAT TECHNIQUES
124
PREDICTION MODEL

FIG. 1A

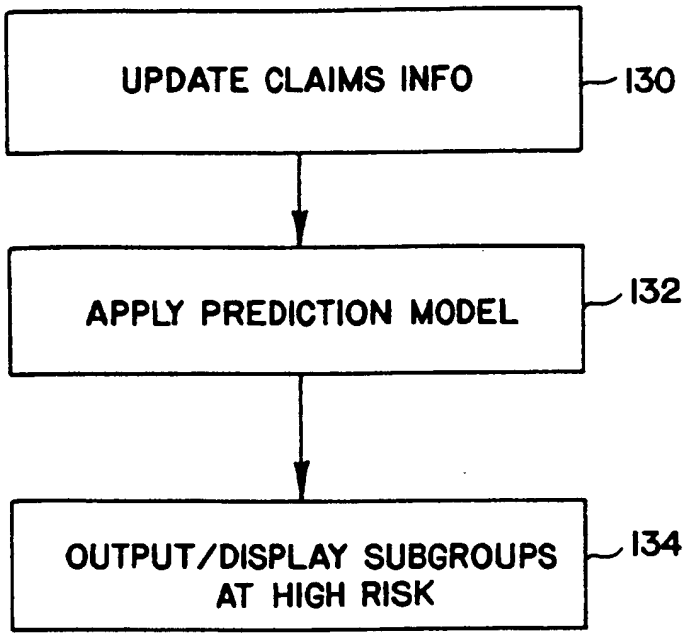
UPDATE CLAIMS INFO
130
APPLY PREDICTION MODEL
132
OUTPUT/DISPLAY SUBGROUPS AT HIGH RISK
134

FIG. 1B

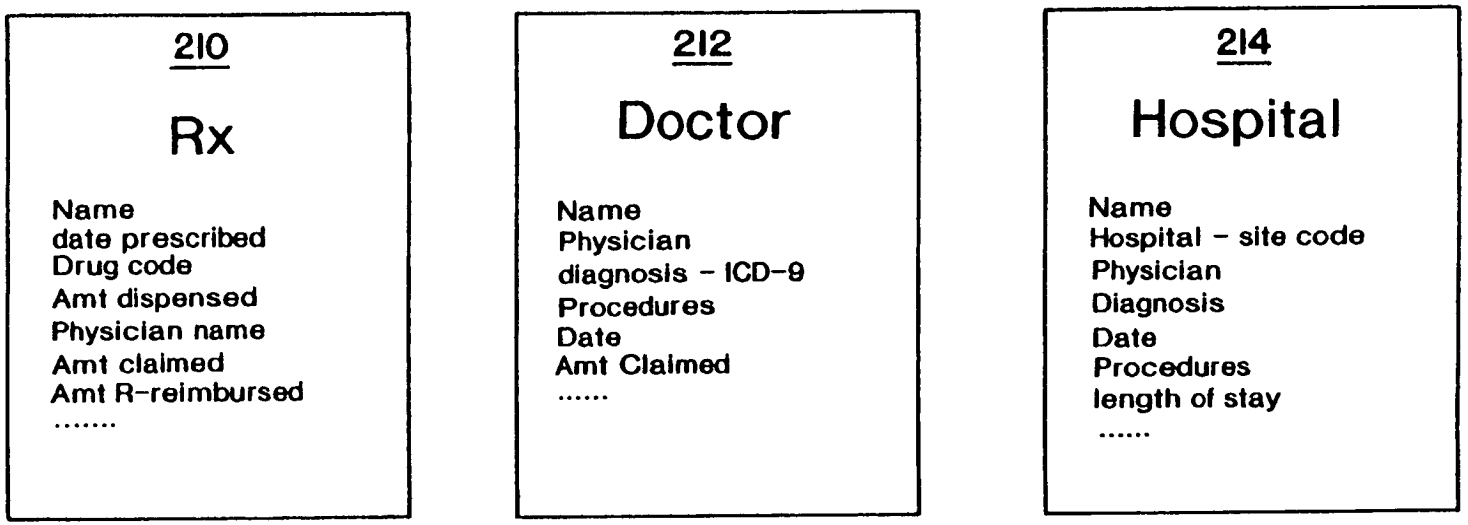

FIG. 2

EP 0 887 759 A1

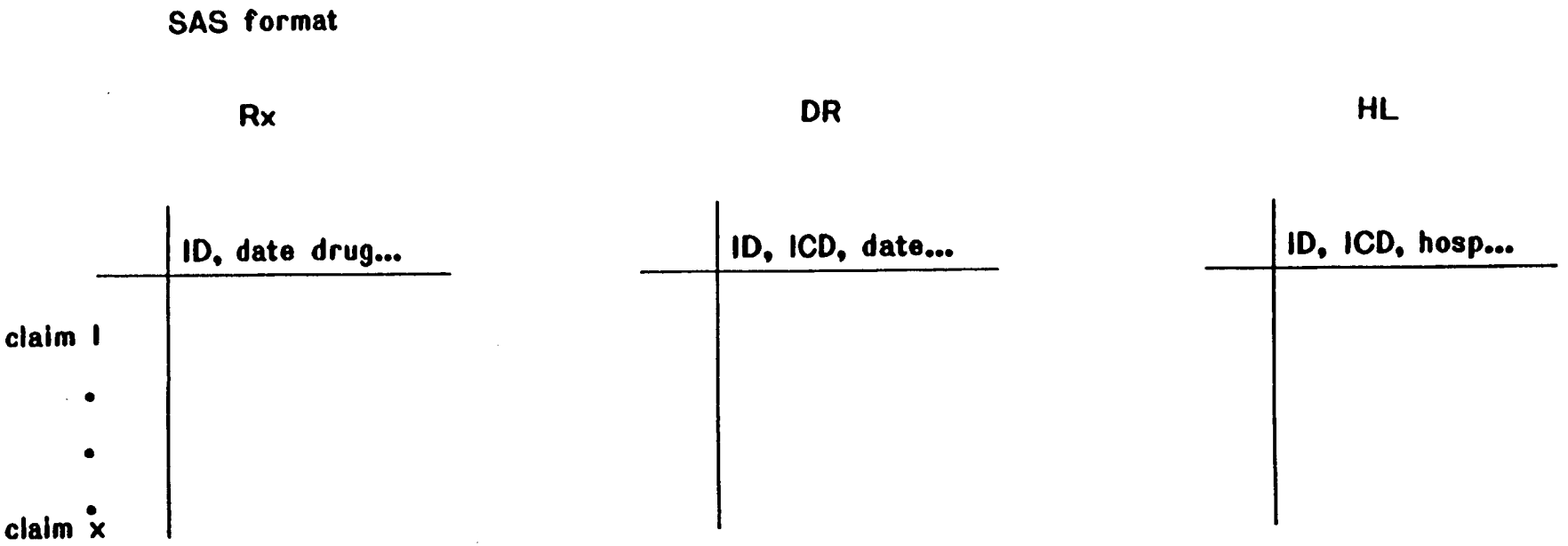

FIG. 3

EP 0 887 759 A1

EVENT	EVENT DATE	NUMBER	COST	CATEGORY INDICATOR
1. CHF ER	date	NA	$	NA
2. CHF Hospitalization	date	LOS	$	NA
3. CHF office visit	date	NA	$	NA
4. IHD ER	date	NA	$	subclass
5. IHD Hospitalization	date	LOS	$	subclass
6. IHD office visit	date	NA	$	subclass
7. Diabetes ER	date	NA	$	NA
8. Diabetes Hospitalization	date	LOS	$	NA
9. Diabetes office visit	date	NA	$	NA
10. Dysrhythmia ER	date	NA	$	NA
11. Dysrhythmia Hospitalization	date	LOS	$	NA
12. Dysrhythmia office visit	date	NA	$	NA
13. Hypertension ER	date	NA	$	NA
14. Hypertension Hospitalization	date	LOS	$	NA
15. Hypertension office visit	date	NA	$	NA
16. Lifestyle ER	date	NA	$	NA
17. Lifestyle Hospitalization	date	LOS	$	NA
18. Lifestyle office visit	date	NA	$	NA
19. Other Hrt Dz ER	date	NA	$	NA
20. Other Hrt DZ Hospitalization	date	LOS	$	NA
21. Other Hrt Dz office visit	date	NA	$	NA
22. Respiratory ER	date	NA	$	NA
23. Respiratory Hospitalization	date	LOS	$	NA
24. Respiratory office visit	date	NA	$	NA
25. Thyrotoxicosis ER	date	NA	$	NA
26. Thyrotoxicosis Hospitalization	date	LOS	$	NA
27. Thyrotoxicosis office visit	date	NA	$	NA
28. Pulmonary Embolism ER	date	NA	$	NA
29. Pulmonary Embolism Hosp.	date	LOS	$	NA
30. Pulmonary Embolism office	date	NA	$	NA
31. Anemia ER	date	NA	$	NA
32. Anemia Hospitalization	date	LOS	$	NA
33. Anemia office visit	date	NA	$	NA
34. Infection ER	date	NA	$	NA
35. Infection Hospitalization	date	LOS	$	NA
36. Infection office visit	date	NA	$	NA
37. Other ER	date of first	# ER in month	$	NA
38. Other Hospitalization	date of first	# hosp in month	$	NA
39. Other office visit	date of first	# OV in month	$	NA
40. Miscellaneous Medical event	date of first	# in month	$	# of CV in month
41. Routine CV Procedures	date of first	# on event date	$	NA
42. Intermediate CV Procedures	date of first	# on event date	$	NA
43. Critical CV Procedures	date of first	# on event date	$	NA
44. CV Surgery	date of first	# on event date	$	NA
45. Rx for ACE inhibitor therapy	date of first	days supply	$	# rx in month
46. Rx for Loop diuretic therapy	date of first	days supply	$	# rx in month
47. Rx for other diuretic therapy	date of first	days supply	$	# rx in month
48. Rx for Digoxin therapy	date of first	days supply	$	# rx in month
49. Rx for Beta Blocker therapy	date of first	days supply	$	# rx in month
50. Rx for Ca Channel Blocker tx	date of first	days supply	$	# rx in month
51. Rx for other CV drug	date of first	days supply	$	# rx in month
52. Rx for non-CV drug	date of first	days supply	$	# rx in month
53. Rx for Na/H20 drug	date of first	days supply	$	# rx in month

FIG. 4

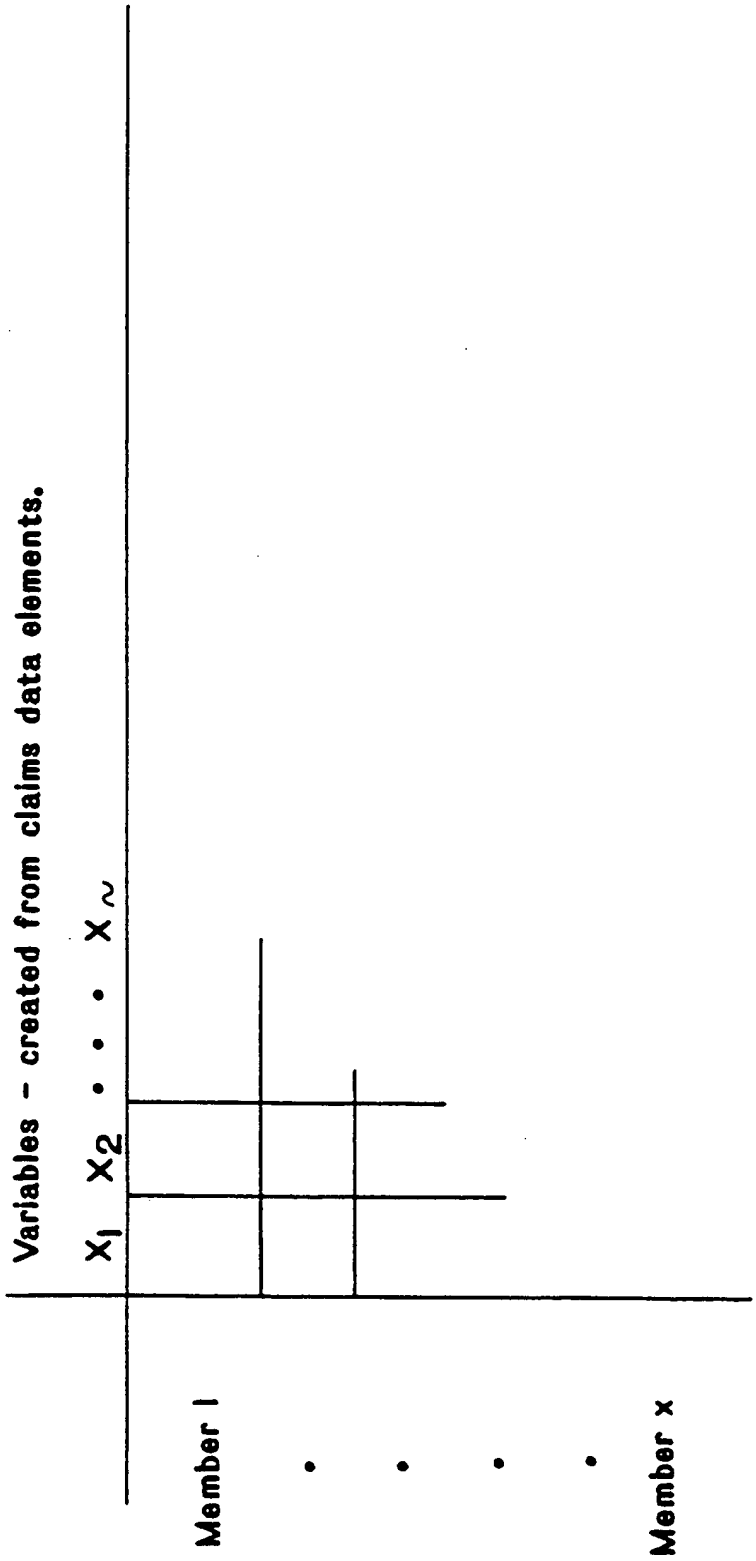
Variables - created from claims data elements.
X1 X2 • • • Xn
Member 1
Member x

FIG. 5

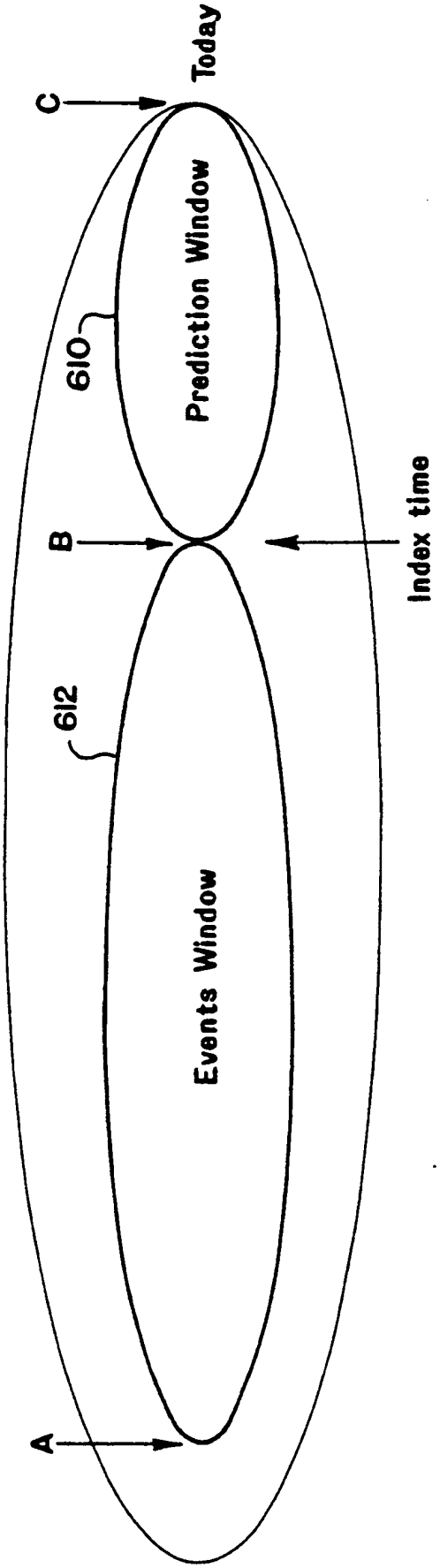
Today
C
Prediction Window
610
B
Index time
612
Events Window
A

FIG. 6A

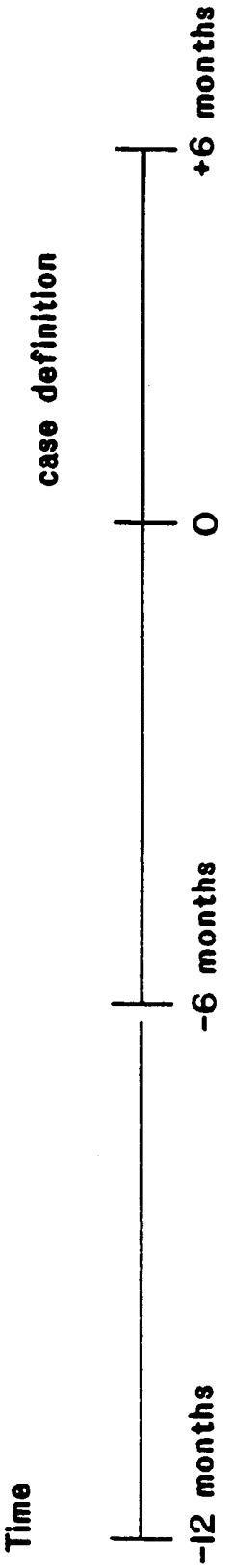
Hospital vs. Hospital
case definition
Time
-12 months
-6 months
0
+6 months
event x 1/e -time

FIG. 6B

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 97 30 4388

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
A	D'HOORE W ET AL: "Risk adjustment in outcome assessment: the Charlson comorbidity index" METHODS OF INFORMATION IN MEDICINE, NOV. 1993, GERMANY, vol. 32, no. 5, ISSN 0026-1270, pages 382-387, XP002055351 * the whole document *	1	G06F19/00
A	KOSTIS W J ET AL: "ESTIMATION OF LONG-TERM VITAL STATUS OF PATIENTS AFTER MYOCARDIAL INFARCTION USING A NEURAL NETWORK BASED ON THE ALOPEX ALGORITHM" PROCEEDINGS OF THE NORTHEAST BIOENGINEERING CONFERENCE, NEWARK, MAR. 18 - 19, 1993, no. CONF. 19, 18 March 1993, LI J K-J;REISMAN S S, page 99/100 XP000399747 * the whole document *	1	
A	DALLE MOLE J ET AL: "A MEDICAL DECISION MAKING SYSTEM FOR PREDICTING THE PROBABILITY OF CARDIAC EVENTS IN PATIENTS WITH RESIDUAL ISCHEMIA AFTER UNCOMPLICATED ACUTE MYOCARDIAL INFARCTION" PROCEEDINGS OF THE COMPUTERS IN CARDIOLOGY MEETING, VENICE, SEPT. 23 - 26, 1991, no. MEETING 18, 23 September 1991, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 229-231, XP000309378 * the whole document *	1	TECHNICAL FIELDS SEARCHED (Int.Cl.6) G06F

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	11 February 1998	Fournier, C

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)